



Reproducible streamflow data analysis on the server or desktop using open source R packages.

David Blodgett

Civil Engineer

U.S. Geological Survey

Office of Water Information

T. Joe Mills

Hydrologist

U.S. Geological Survey

Colorado Water Science Center

Competing Demands for Water Resources



Credit: Erik Eckles, TrekEarth

EflowStats: An Implementation of Ecological Flows Assessment Methods.

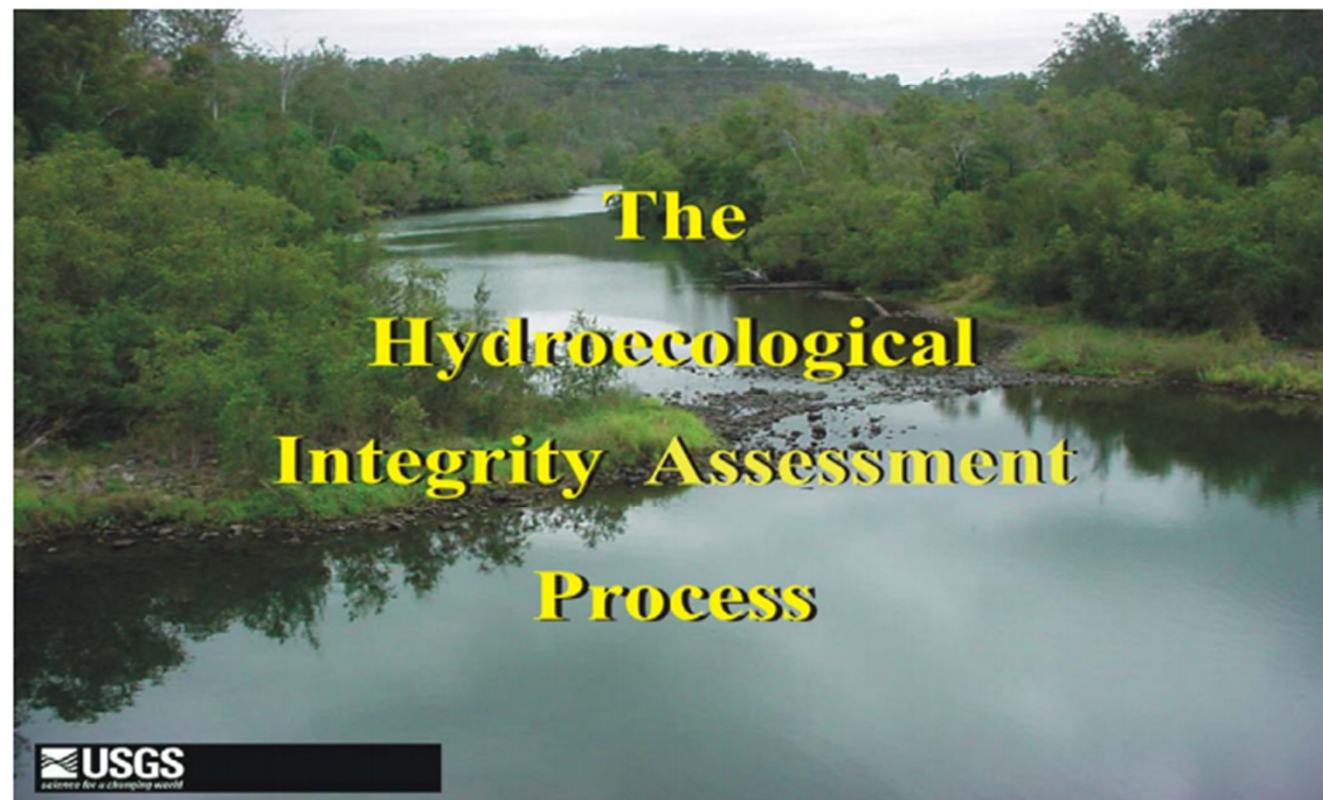
Background

Hydrologic Integrity Assessment Process

- Hydrologic Index Tool
- Hydrologic Assessment Tool
 - Classify streams into groups
 - Identify set of significant indices
 - Classify study stream(s)
 - Establish hydrologic baselines, environmental flow standards, or evaluate proposed changes.

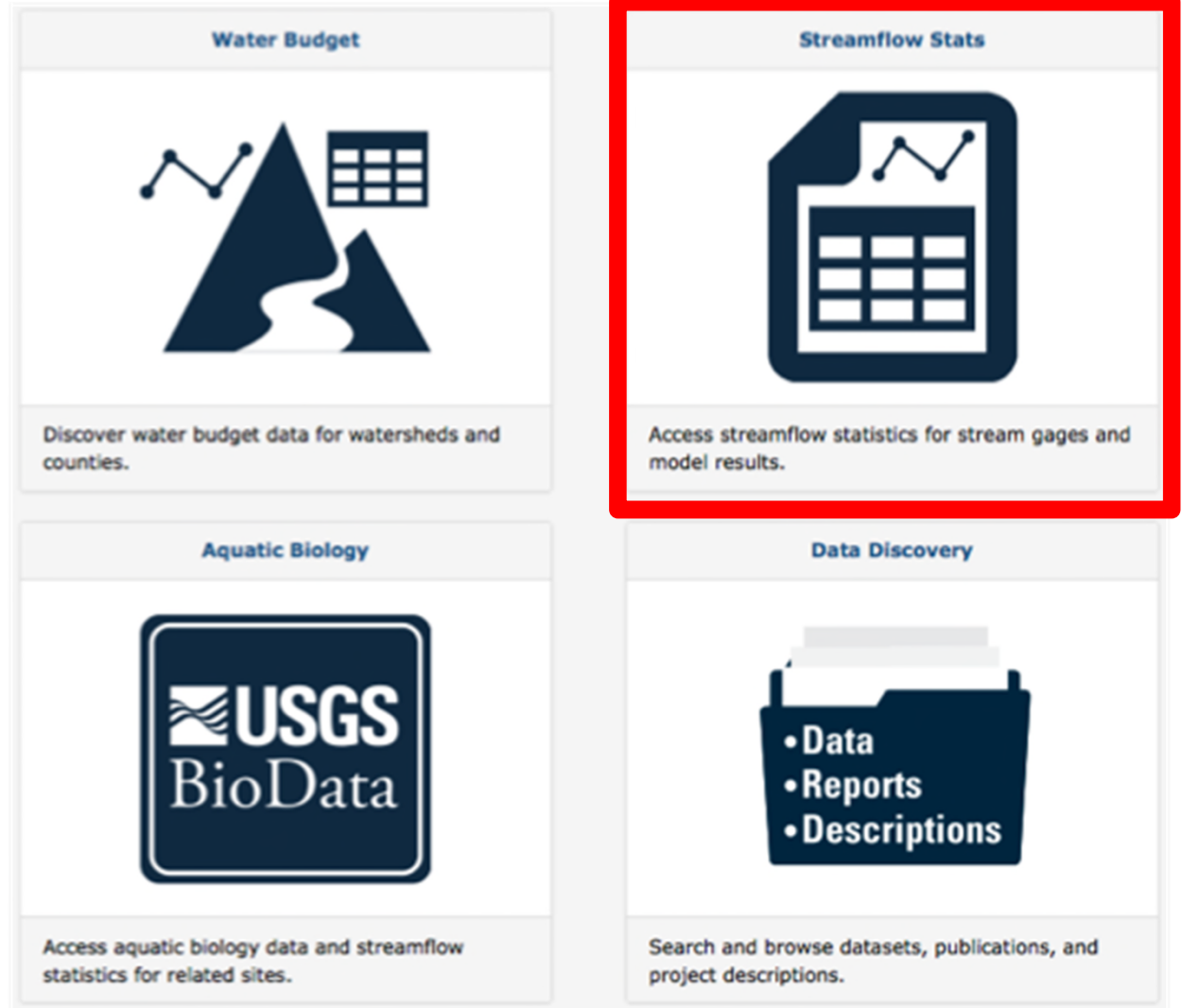
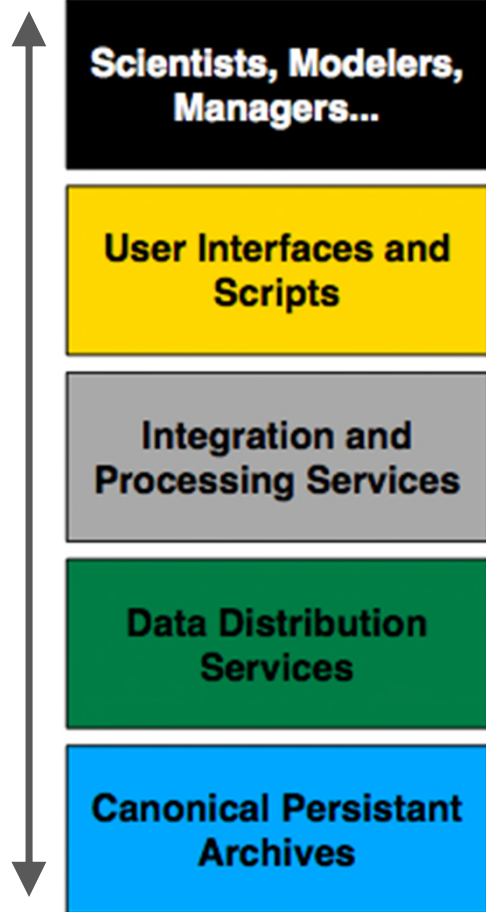
Users' Manual for the Hydroecological Integrity Assessment Process Software (including the New Jersey Assessment Tools)

By James A. Henriksen, John Heasley, Jonathan G. Kennen, and Steven Nieswand

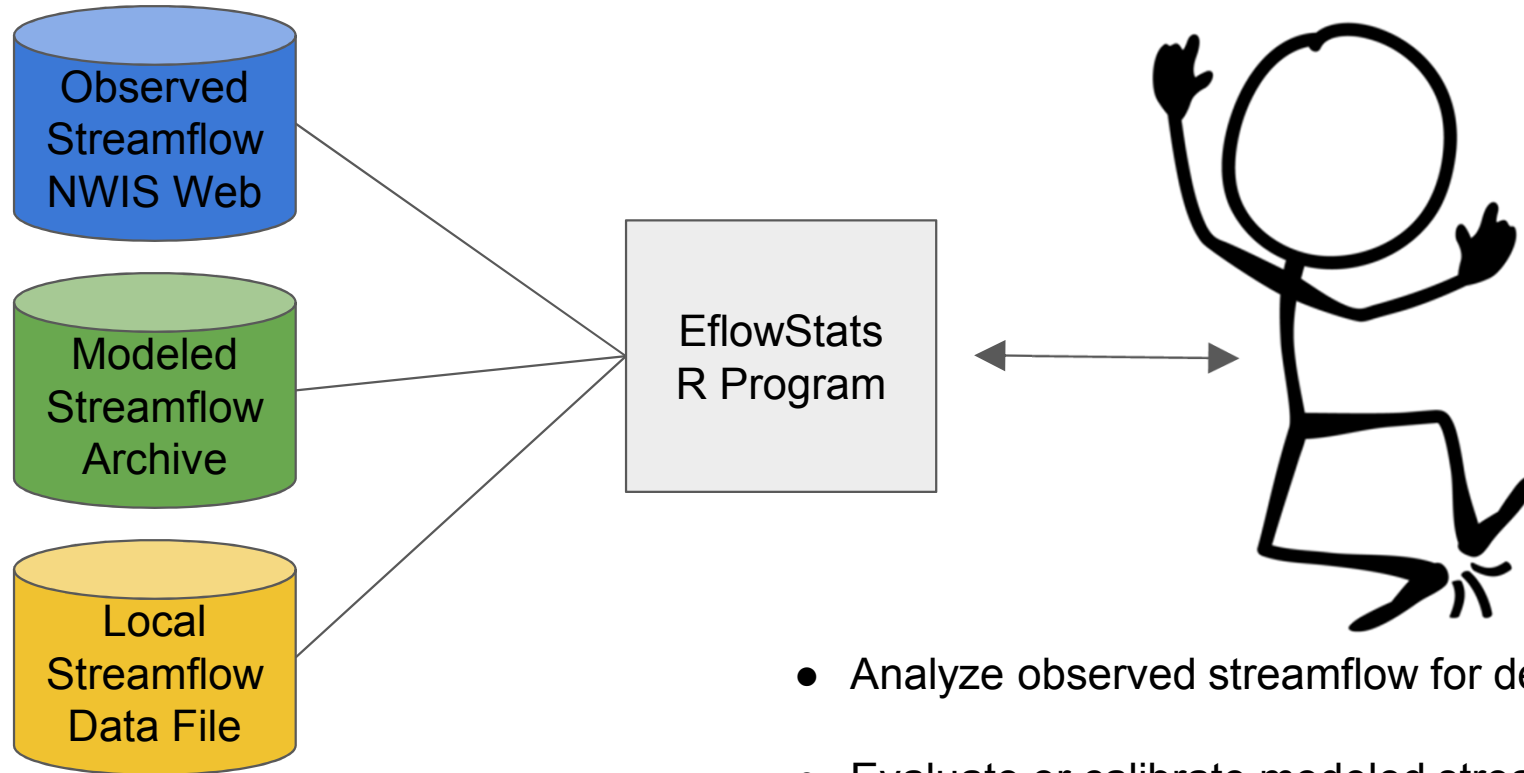


Open File Report 2006-1093

Data Resources Portal: <http://cida.usgs.gov/nwc/>



Data Access from Multiple Sources



- Analyze observed streamflow for decision making.
- Evaluate or calibrate modeled streamflow to reproduce the selected flow metrics.

Package functionality via desktop or web

R ON DESKTOP

Programmers and researchers can access particular statistic functions.



Decision makers and engineers can access suites of statistics and other summary functions.



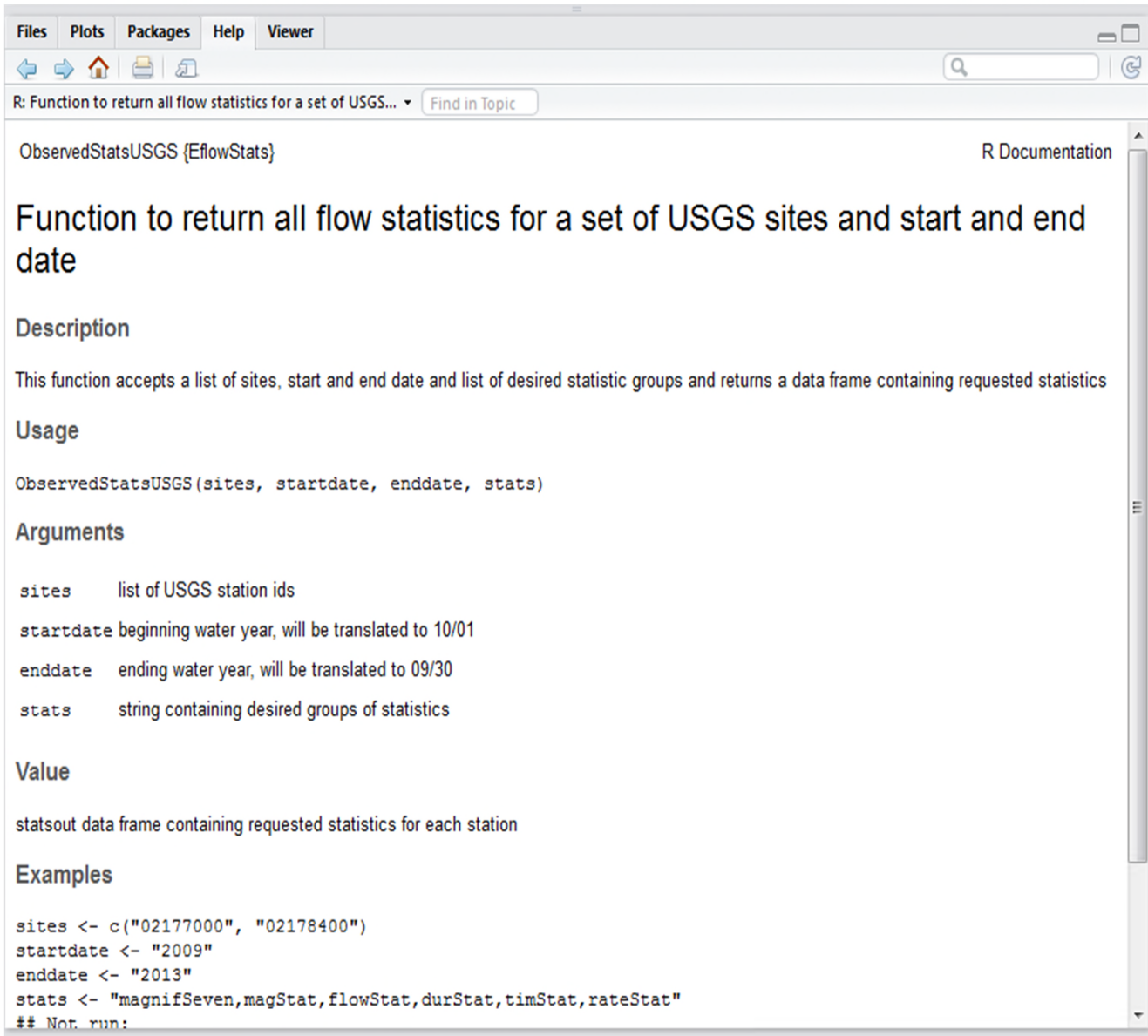
R BEHIND WEBPAGE

Any user can access the functionality via a web page for exploratory or curiosity driven investigations.



EflowStats on the desktop

- Run EflowStats in the R or RStudio environment
- Scripting for reproducible analysis
- All functions cataloged
- Help files with examples

A screenshot of the RStudio Help window. The title bar shows "Files", "Plots", "Packages", "Help", and "Viewer". The search bar at the top contains "R: Function to return all flow statistics for a set of USGS..." and a "Find in Topic" button. The main content area displays the documentation for the `ObservedStatsUSGS` function. It includes the function name, a brief description, a detailed description, usage instructions, arguments, value, and examples. The examples section shows R code for setting sites, start/end dates, and statistics groups.

ObservedStatsUSGS {EflowStats} R Documentation

Function to return all flow statistics for a set of USGS sites and start and end date

Description

This function accepts a list of sites, start and end date and list of desired statistic groups and returns a data frame containing requested statistics

Usage

```
ObservedStatsUSGS(sites, startdate, enddate, stats)
```

Arguments

`sites` list of USGS station ids

`startdate` beginning water year, will be translated to 10/01

`enddate` ending water year, will be translated to 09/30

`stats` string containing desired groups of statistics

Value

statsout data frame containing requested statistics for each station

Examples

```
sites <- c("02177000", "02178400")
startdate <- "2009"
enddate <- "2013"
stats <- "magnifSeven,magStat,flowStat,durStat,timStat,rateStat"
## Not run:
```

EflowStats on the desktop

Low-level functions for custom workflows

```
#Specify station and parameters
##Select USGS site numbers
sites <- c("02177000", "02178400")

##Set date range
startdate <- "2009-01-01"
enddate <- "2013-01-01"

#Get Data
data <- getDataUSGS(sites,startdate,enddate)

#Calculate Statistics
fh1.2(data[[1]])
```

```
Console ~/USGS-R/EflowStats/ ↗
>
> #Calculate statistics
> fh1.2(data[[1]])
+

> #Specify station and parameters
> ##Select USGS site numbers
> sites <- c("02177000", "02178400")
>
> ##Set date range
> startdate <- "2009-01-01"
> enddate <- "2013-01-01"
>
> #Get Data
> data <- getDataUSGS(sites,startdate,enddate)
get_obsdata run on x_obs for site 02177000 1096
get_obsdata run on x_obs for site 02178400 1096
>
> #Calculate statistics
> fh1.2(data[[1]])
$fh1
[1] 11.33

$fh2
[1] 10.19

> |
```

EflowStats on the desktop

High-level wrapper functions for standard tasks

```
#Specify station and parameters
##Select USGS site numbers
sites <- c("02177000", "02178400")

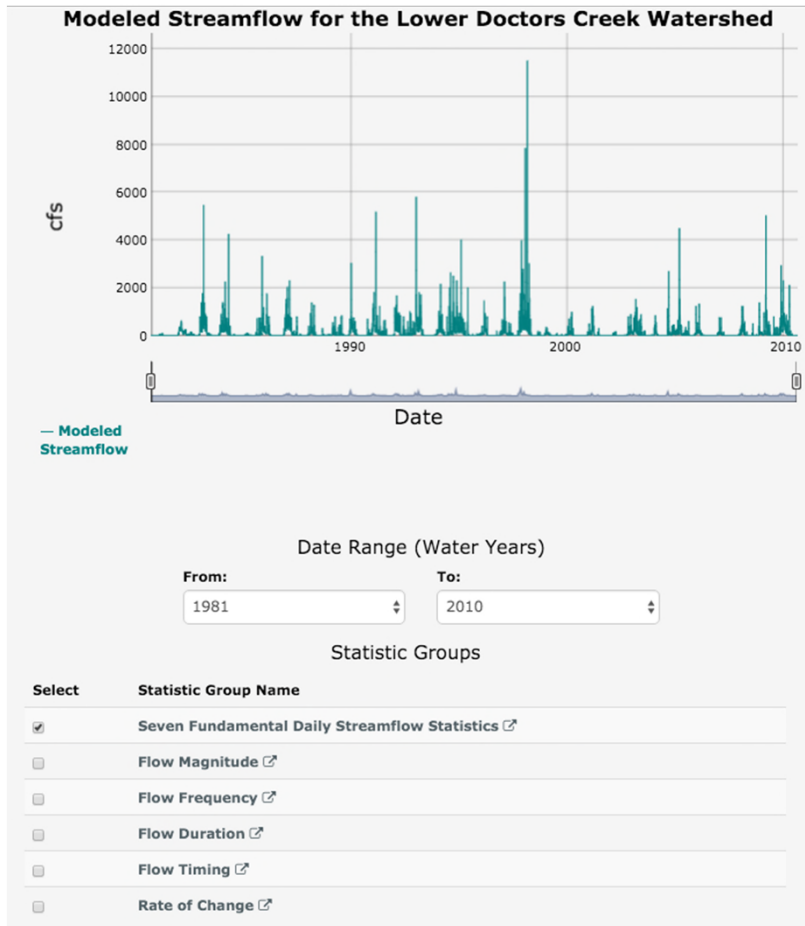
##Set date range
startdate <- "2009"
enddate <- "2013"

##Specify statistics
stats <- "magnifSeven,magStat,flowStat,durStat,timStat,rateStat"

#Calculate statistics
usgsStats <- ObservedStatsUSGS(sites,startdate,enddate,stats)
```

	site_no	min_date	max_date	lam1Obs	tau2Obs	tau3Obs	tau4Obs	ar1Obs	amplitudeObs	phaseObs	ma1_mean_disc	ma2_median_disc	ma3_mean_annual_var	ma4	ma5_skew
1	02177000	2009-10-01	2013-09-30	697.04	0.40	0.31	0.15	0.55	0.45	-1.15	697.04	531.0	71.77	64.78	1.31
2	02178400	2009-10-01	2012-09-30	170.91	0.38	0.35	0.23	0.49	0.61	-0.82	170.91	136.5	83.46	61.20	1.25

EflowStats on Data Resources Portal



Observed and
Modeled Streamflow
Same Statistics

Select a subset of the time series for which you would like to calculate various statistics.

Gage ID

02198000

Access data for this site from [NWIS-Web](#) [↗](#)

Gage Name

BRIER CREEK AT MILLHAVEN, GA

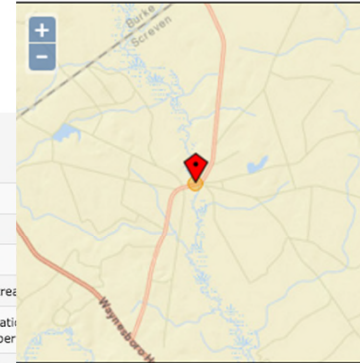
Gage Watershed Drainage Area (km²)

1675.712

Period of Record

1937-04-14

2016-04-13



Name	Value	Description
site_no	030701060404	
min_date	1980-10-01	
max_date	2010-09-30	
lam1Obs	140.72	Mean of the daily streamflow for the period of record.
tau2Obs	0.76	Coefficient of L-variability of the daily streamflow for the period of record.
tau3Obs	0.62	L-skewness (Hosking and Wallis, 1994, eqn. 2.60 with r equal to 3) of the daily streamflow for the period of record.
tau4Obs	0.38	L-kurtosis (Hosking and Wallis, 1994, eqn. 2.60 with r equal to 4) of the daily streamflow for the period of record.
ar1Obs	-0.02	Autoregressive lag-one correlation coefficient of the daily streamflow for the period of record. First, the streamflow time series is deseasonalized by subtracting the long-term monthly mean streamflow from each of the days in the respective month and then standardized to have a zero mean and unit variance.
amplitudeObs	0.47	Amplitude of the seasonal signal of the daily streamflow for the period of record. Streamflow is first standardized to have a zero mean and unit variance, a regression of streamflow versus $\sin(2\pi y)$ and $\cos(2\pi y)$ is performed and amplitude is then computed from the regression coefficients as given by Warner (1998).
phaseObs	-0.82	Phase shift of the seasonal signal of the daily streamflow for the period of record. Streamflow is first standardized to have a zero mean and unit variance, a regression of streamflow versus $\sin(2\pi y)$ and $\cos(2\pi y)$ is performed and phase shift is then computed from the regression coefficients as given by Warner (1998).
comment		

Development Distribution



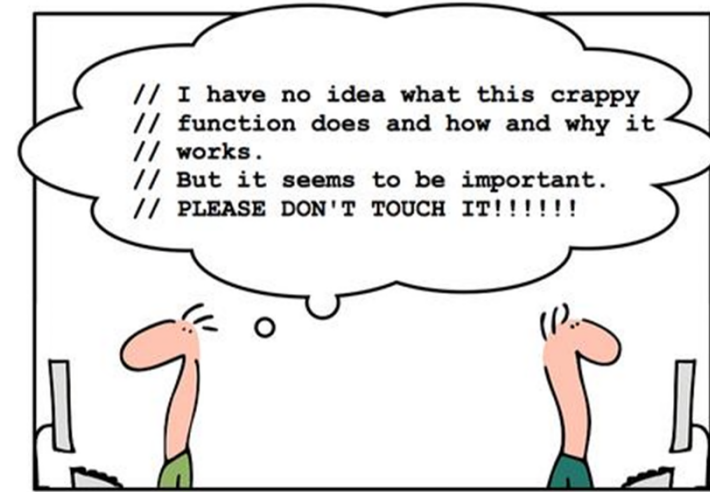
GRAN

Geological Survey R Archive Network



Consistent best practices among USGS scientists

- All code must have an active USGS maintainer listed in the package
- Code must be maintained on GitHub
- Issue tracking in GitHub
- Continuous integration strongly encouraged



Calculates Hydrologic Indicator stats and fundamental properties of daily streamflow for a given set of data

281 commits

1 branch

7 releases

4 contributors

Branch: master

New pull request

New file

Upload files

Find file

HTTPS

https://github.com/USGS-

Download ZIP

tmills-usgs Merge pull request #38 from tmills-usgs/master

Latest commit bf45ca2 on Feb 4

R

Replaced sample data with new dataset contain years lwth no low flow ...

2 months ago

data

Switched back to old sample data because all testthat tests now fail ...

2 months ago

inst

Fixed the site ids.

a year ago

man

Updated getDrainageArea

a year ago

tests

Started frequency function tests

5 months ago

vignettes

Minor editorial changes to vignette

7 months ago

.Rbuildignore

appveyor

6 months ago

.gitignore

added readme to gitignore

6 months ago

.travis.yml

Added slack integration

2 months ago

DESCRIPTION

Update DESCRIPTION

9 months ago

LICENSE.note

Renamed file.

a year ago

NAMESPACE

Taking out XML and RCurl (imports from dataRetrieval).

a year ago

README.md

Update README.md

6 months ago

appveyor.yml

appveyor

6 months ago

README.md

EflowStats

Calculates a suite of ecological flow statistics and fundamental properties of daily streamflow for a given set of data.

The `eflowstats` package was created to simplify the process of generating hydrologic indicator statistics using daily streamflow records. It has been specifically designed to work seamlessly with U.S. Geological Survey (USGS) National Water Information System (NWIS) data. This package is intended to be an update of the previously existing USGS National Hydrologic Assessment Tool (NAHAT) program with additional statistics previously published by Archfield et al (Archfield).

Installation

To install this package use the following code:


```
install.packages("EflowStats", repos=c("http://owi.usgs.gov/R", "http://cran.us.r-project.org"))
```

Linux Tests: build passing




Windows Tests: build passing

References












 This repository Search

Pull requestsIssuesGist

USGS-R / EflowStats

 Unwatch 10  Star 4  Fork 11

 Code  Issues 9  Pull requests 0  Wiki  Pulse  Graphs



Filters

LabelsMilestones



New issue

☐ 9 Open ✓ 12 Closed



Author ▾ Labels ▾ Milestones ▾ Assignee ▾ Sort ▾

☐  **Loops in many functions need to be eliminated** Performance  0



#43 opened 2 days ago by tmills-usgs

☐  **Assess performance, spawn new github issues to address issues found.** enhancement  1



#41 opened 20 days ago by dbloodgett-usgs

☐  **Match all function names to .r file name**  0



#36 opened on Feb 4 by tmills-usgs

☐  **Clarify and possibly add constraints for POR of diffstats** enhancement  3



#25 opened on Jul 9, 2015 by dbloodgett-usgs

☐  **allow runs with less than complete water years of data** enhancement  1



#24 opened on Jul 9, 2015 by dbloodgett-usgs

☐  **Add comment to output to tell user which water years were removed** enhancement  0



#23 opened on Jul 9, 2015 by dbloodgett-usgs

☐  **Ability to define water year other than Oct-Oct** enhancement  1


#22 opened on Jul 9, 2015 by dbloodgett-usgs

☐  **Review function documentation.** enhancement  1

#19 opened on Jul 9, 2015 by dbloodgett-usgs

☐  **characterization testing**  3

#18 opened on Jul 7, 2015 by dbloodgett-usgs

 ProTip! Mix and match filters to narrow down what you're looking for.



Assess performance, spawn new github issues to address issues found. #41

 **Open** **dblodgett-usgs** opened this issue 20 days ago · 1 comment



dblodgett-usgs commented 20 days ago

USGS-R member



We've gotten a request to improve the performance of the package. Working on getting more specifics, but a basic review of the code with an eye on performance is likely in order. I'm guessing there's a pretty easy to identify list of things that will help speed up processing.



tmills-usgs was assigned by **dblodgett-usgs** 20 days ago



dblodgett-usgs added the **enhancement** label 20 days ago



tmills-usgs commented 20 days ago

USGS-R member



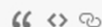
I will take a look but I agree that there are some low hanging fruit that would probably streamline the code substantially



Write

Preview

AA B i



Leave a comment

Attach files by dragging & dropping, [selecting them](#), or pasting from the clipboard.

 Styling with Markdown is supported

Close issue

Comment



```
376 * checking package directory ... OK
377 * checking DESCRIPTION meta-information ... OK
378 * checking top-level files ... OK
379 * checking for left-over files ... OK
380 * checking index information ... OK
381 * checking package subdirectories ... OK
382 * checking R files for non-ASCII characters ... OK
383 * checking R files for syntax errors ... OK
384 * checking whether the package can be loaded ... OK
385 * checking whether the package can be loaded with stated dependencies ... OK
386 * checking whether the package can be unloaded cleanly ... OK
387 * checking whether the namespace can be loaded with stated dependencies ... OK
388 * checking whether the namespace can be unloaded cleanly ... OK
389 * checking loading without being on the library search path ... OK
390 * checking dependencies in R code ... OK
391 * checking S3 generic/method consistency ... OK
392 * checking replacement functions ... OK
393 * checking foreign function calls ... OK
394 * checking R code for possible problems ... OK
395 * checking Rd files ... OK
396 * checking Rd metadata ... OK
397 * checking Rd line widths ... OK
398 * checking Rd cross-references ... OK
399 * checking for missing documentation entries ... OK
400 * checking for code/documentation mismatches ... OK
401 * checking Rd \usage sections ... OK
402 * checking Rd contents ... OK
403 * checking for unstated dependencies in examples ... OK
404 * checking contents of 'data' directory ... OK
405 * checking data for non-ASCII characters ... OK
406 * checking data for ASCII and uncompressed saves ... OK
407 WARNING
408 'qdof' is needed for checks on size reduction of PDFs
409 * checking installed files from 'inst/doc' ... OK
410 * checking for old-style vignette sources ... NOTE
411 Vignette sources only in 'inst/doc':
412   'EflowStats.Rnw'
413 A 'vignettes' directory is required as from R 3.1.0
414 and these will not be indexed nor checked
415 * checking examples ... OK
416 Examples with CPU or elapsed time > 5s
417           user system elapsed
418 getDatalocal    12.00   0.00   12.00
419 ObservedStatsOther  6.09   0.02   6.36
420 * checking for unstated dependencies in 'tests' ... OK
421 * checking tests ...
422   Running 'testthat.R'
423   OK
424 * DONE
425
426 Status: 1 WARNING, 2 NOTES
427 See
428   'C:/projects/eflowstats/EflowStats.Rcheck/00check.log'
429 for details.
430
431
432 + [[ -n '' ]]
433 + [[ -n '' ]]
434 Packaging artifacts...OK
435 Uploading artifact EflowStats.Rcheck\00check.log (3 KB)...OK
436 Uploading artifact EflowStats.Rcheck\00install.out (357 bytes)...OK
437 Uploading artifact EflowStats.Rcheck\EflowStats-Ex.Rout (104.1 KB)...OK
438 Uploading artifact EflowStats.Rcheck\tests\testthat.Rout (8.6 KB)...OK
439 Uploading artifact eflowstats_4.1.0.tar.gz (11.1 MB)...OK
440 Uploading artifact EflowStats_4.1.0.zip (1.2 MB)...OK
441 Build success
```



Installation

To install this package use the following code:

```
install.packages("EflowStats",repos=c("http://ow1.usgs.gov/R","http://cran.us.r-project.org"))
```

Linux Tests: build passing

Windows Tests: build passing

References

WikiPulseGraphs

fundamental properties of daily streamflow for a given set of data

1 branch7 releases4 contributors

fileUpload filesFind fileHTTPShttps://github.com/USGS-Download ZIP

Latest commit bf45ca2 on Feb 4

with new dataset contain years iwth no low flow ...2 months ago

ample data because all testthat tests now fail ...2 months ago

a year ago

a year ago

5 months ago

7 months ago

6 months ago

6 months ago

2 months ago

9 months ago

a year ago

a year ago

6 months ago

6 months ago

Statistics and fundamental properties of daily streamflow for a given set of data.

To simplify the process of generating hydrologic indicator statistics using daily cally designed to work seamlessly with U.S. Geological Survey (USGS) National This package is intended to be an update of the previously existing USGS National program, with additional statistics previously published by Archfield et al (Archfield).

Linux Tests: build error

Windows Tests: build passing

Scalable, stable, and collaborative research applications

Scalable

Low-level functions

Customized workflows



High-level functions

Standard outputs



Web deployment

Dissemination to public

Stable

- Employ minimum set of best practices
- Code management and version control
 - Git and GitHub
- Continuous Integration
 - Modify code != different results

Collaborative

- Consistent format for GRAN R packages
- Scientist focuses on code and analysis not framework design
- Github code availability facilitates collaboration

Questions???

